Project Acronym:	**VICINITY**
Project Full Title:	**Open virtual neighbourhood network to connect intelligent buildings and smart objects**
Grant Agreement:	**688467**
Project Duration:	**48 months (01/01/2016 - 31/12/2019)**

Deliverable D9.2

Data Management Plan, first version

Work Package:	**WP9 – Dissemination of Results & Exploitation**
Task(s):	**T9.2 – Data Management Plan**
Lead Beneficiary:	**CERTH**
Due Date:	**30 June 2016 (M6)**
Submission Date:	**30 June 2016 (M6)**
Deliverable Status:	**Final**
Deliverable Type:	**R**
Dissemination Level:	**PU**
File Name:	**VICINITY_D9.2_DMP_1_2.pdf**

VICINITY Consortium

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°688467

No	Beneficiary		Country
1.	TU Kaiserslautern (Coordinator)	UNIKL	Germany
2.	ATOS SPAIN SA	ATOS	Spain
3.	Centre for Research and Technology Hellas	CERTH	Greece
4.	Aalborg University	AAU	Denmark
5.	GORENJE GOSPODINJSKI APARATI D.D.	GRN	Slovenia
6.	Hellenic Telecommunications Organization S.A.	OTE	Greece
7.	bAvenir s.r.o.	BVR	Slovakia
8.	Climate Associates Ltd	CAL	United Kingdom
9.	InterSoft A.S.	IS	Slovakia
10.	Universidad Politécnica de Madrid	UPM	Spain
11.	Gnomon Informatics S.A.	GNOMON	Greece
12.	Tiny Mesh AS	TINYM	Norway
13.	HAFENSTROM AS	HITS	Norway
14.	Enercoutim – Associação Empresarial de Energia Solar de Alcoutim	ENERC	Portugal
15.	Municipality of Pylaia-Hortiatis	MPH	Greece

Disclaimer

This document reflects only the author's views and the European Union is not liable for any use that may be made of the information contained therein.

Authors List

Leading Author (Editor)			
Surname	**First Name**	**Beneficiary**	**Contact email**
Sveen	Flemming	HITS	flsveen@online.no

Co-authors (in alphabetic order)				
No	**Surname**	**First Name**	**Beneficiary**	**Contact email**
1.	Guerrero	Josep M.	AAU	joz@et.aau.dk
2.	Hovstø	Asbjørn	HITS	hovsto@online.no
3.	Kaggelides	Konstantinos	GNOMON	k.kaggelides@gnomon.com.gr
4.	Rico	Juan	ATOS	juan.rico@atos.net
5.	Samovich	Natalie	ENERC	n.samovich@enercoutim.eu
6.	Tryferidis	Athanasios	CERTH	thanasic@iti.gr

Reviewers List

List of Reviewers (in alphabetic order)				
No	**Surname**	**First Name**	**Beneficiary**	**Contact email**
1.	Dickerson	Keith	CAL	Keith.dickerson@mac.com

Revision Control

Version	Date	Status	Modifications made by
0.1	22. April 2016	Initial Draft	Sveen (HITS)
0.2	9. May 2016	First Draft formatted with contributions received	Tryferidis (CERTH), Rico (ATOS), Sveen (HITS)
0.3	16. May 2016	Deliverable version for final review by partners	Sveen (HITS)
0.4	23. May 2016	Final improvements	Sveen (HITS)
0.5	3. June 2016	Deliverable version uploaded for Quality Check	Sveen (HITS)
0.6	21. June 2016	Quality Check	Sveen (HITS)
0.7	22. June 2016	Final Draft reviewed	Sveen (HITS)
0.8	28. June 2016	Final datasets included	Kaggelides (GNOMON), Samovich (ENERC), Tryferidis (CERTH), Rico (ATOS), Hovstø (HITS, Sveen (HITS)
0.9	29. June 2016	Final improvements	Tryferidis (CERTH), Sveen (HITS)
1.0	**30. June 2016**	**Submission to the EC**	**Sveen (HITS)**

List of Definitions & Abbreviations

Abbr.	Definition (A-G)	Abbr.	Definition (H-Q)	Abbr.	Definition (R-X)
AI	Artificial Intelligence	HTTP	Hypertext Transport Protocol	RDF	Resource Description Framework
API	Application Programming Interface	ICT	Information & Communication Technologies	RES	Renewable Energy Resources
ASN	Abstract Syntax Notation	IEQ	Indoor Environmental Quality	RFID	Radio Frequency Identification
CA	Consortium Agreement	IoT	Internet of Things	SG	Study Group
CL	Classified	IP	Internet Protocol	SM	Scientific Manager
CO	Confidential	IPR	Intellectual Property Right	SME	Small Medium Enterprise
DER	Distributed Energy Resources	IR	Infrared	SOA	Service Oriented Architecture
DG	Distribution Grid	ITU	International Telecommunication Union	SSL	Secure Socket Layer
DL	Description logic	M2M	Machine to Machine	SSN	Social Security Number / Semantic Sensor Network
DSM	Digital Single Market	MQTT	Message Queuing Telemetry Transport	SW	Software
DSM	Demand Side Management	NFC	Near Field Communication	TCP	Transmission Control Protocol
DSO	Distribution System Operator	OSG	Open Geospatial Consortium	TSO	Transmission System Operator
EC	European Commission	OS	Operating System	UA	Unified Architecture
ESCO	Energy Service COmpany	OWL	Web Ontology Language	UDP	User Datagram Protocol

ESO	European Standards Organization	PC	Project Coordinator	URL	Uniform Resource Locator
ETSI	European Telecommunications Standards Institute	PO	Project Officer	UUID	Universally unique identifier
EU	European Union	PU	Public	VNM	Vicinity Neighbourhood Manager
GPS	Geographic Positioning System	QA	Quality Assurance	WP	Work Package
GSM	Global System for Mobile communications			XML	EXtensible Markup Language

List of figures:

Table of Contents

1. Executive Summary

> *«The VICINITY project will build and demonstrate a bottom-up ecosystem of decentralised interoperability of IoT infrastructures called virtual neighborhood, where users can share the access to their smart objects without losing the control over them.»*

The present document is a deliverable “D9.2 – Data Management Plan” of the VICINITY project of the VICINITY project (Grant Agreement No.: 688467), funded by the European Commission’s Directorate-General for Research and Innovation (DG RTD), under its Horizon 2020 Research and Innovation Programme (H2020).

The VICINITY Consortium has identified several areas that needs to be addressed; Protocol interoperability, identification tokens, encryption keys, data formats and packet size. Also, several issues are related to latency, bandwidth and general architecture.

VICINITYs activities will involve human participants, as some of the pilots will be conducted in real homes with actual residents. For some of the activities to be carried out by the project, it may be necessary to collect basic personal data (e.g. name, background, contact details), even though the project will avoid collecting such data unless necessary. Such data will be protected in accordance with the EU's Data Protection Directive 95/46/EC[1] of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. National and local legislations applicable to the project will also be strictly applied (full list described in annex 2: ethics and security).

All personal data, or data directly related to the residents, will first be collected when the project has received a signed informed consent form from the subjects participating.

This is the first version of the project Data Management Plan (DMP). It contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The purpose of the Data Management Plan (is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project. The DMP is not a fixed document, but will evolve during the lifespan of the project.

The datasets referred to in this document are drafted during the first project stages (completed 30th of June 2016) of the project. The document can only reflect the intentions of the project partners toward developing the overall project’s datasets. The second revision (D9.3) will be prepared by 31st December 2017, and the third (D9.4) will be ready by 31st December 2019. This follows the H2020 guidelines on Data Management Plans, and as stated in the Grant Agreement 688467.

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en

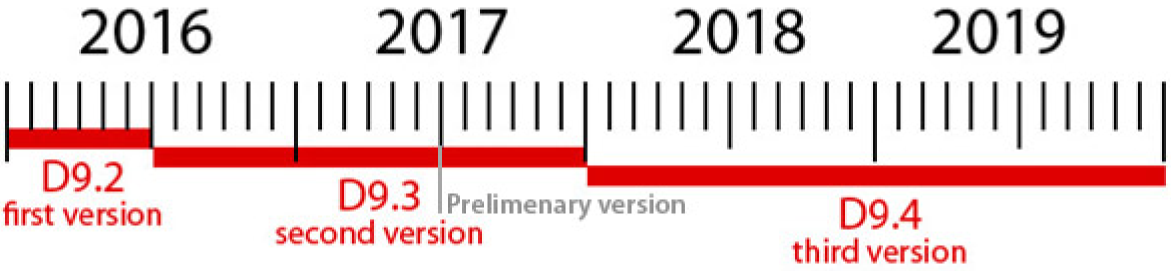

Figure 1: Data Management Plan – deliverables 2016 – 2019

Note: In order to assist the official project review process by the commission for the first project period (M1-M24), a preliminary version of the updated DMP of D9.3 can be further delivered prior to M24 (December 2017), in order to be enable a better assessment of the progress of the Data Management in the project by the reviewers.

As the project progresses and results start to arrive, the datasets will be elaborated on. The detailed descriptions of all the specific datasets that have been collected will be described, made available under the relevant Data Management framework.

2. Introduction

The purpose of the Data Management Plan (DMP) deliverable is to provide relevant information concerning the data that will be collected and used by the partners of the project VICINITY. The project aims to develop a solution defined as "Interoperability as a Service" which will be a part of the VICINITY open gateway. In order to achieve this, a platform for harvesting, converting and sharing data from IoT units has to be implemented on the service layer of the network.

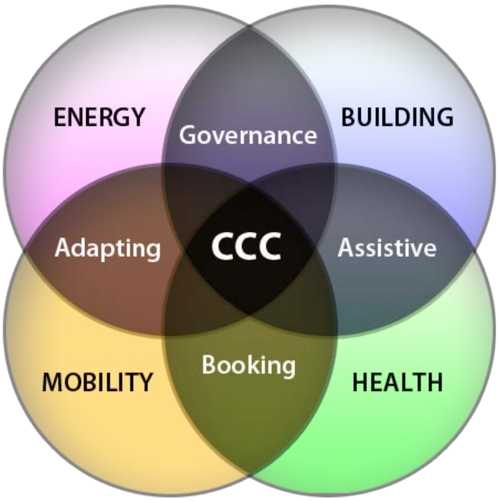

Figure 2: Domains and some of the functionalities the DMP has to cover

This goal entails the need for good documentation and implementation of descriptors, lookup-tables, privacy settings and intelligent conversion of data formats. The strength of having a cloud-based gateway is that it should be relatively simple to upgrade with new specifications and implement conversion, distribution and privacy strategies. In particular, the privacy part is considered an important aspect of the project, as VICINITY needs to follow and adhere to strict privacy policies. It will also be necessary to focus on possible ethical issues and access restrictions regarding personal data so that no regulations on sensitive information are violated.

The datasets collected will belong to four main domains; smart energy, mobility, smart home and eHealth. There exists several standards and guidelines the project needs to be aware within each of these fields. There are a number of different vendors and disciplines involved – and much of the information that is available only exists in proprietary data formats. For this reason, VICINITY will target IoT units that follows the specifications defined by oneM2M consortium, ETSI standardization group and international groups and committees.

The DMP will be refined in subsequent revisions of the present deliverables. This first version of the DMP mainly depicts what the project currently (M6) expects of the direction regarding the collection of the data.

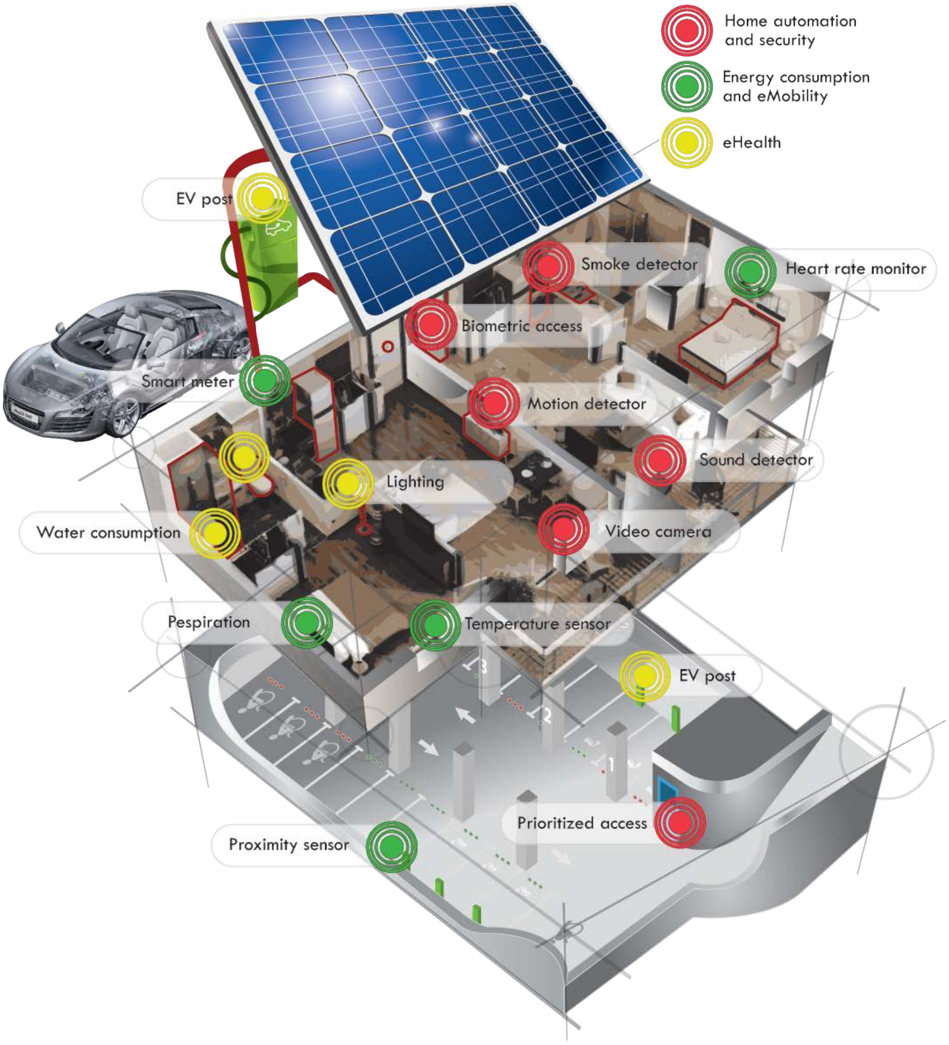

Figure 3: Example of potential data points in use cases that generate data.

3. General Principles

3.1.Participation in the Pilot on Open Research Data

VICINITY participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 programme. The consortium believes firmly in the concepts of open science, and the large potential benefits the European innovation and economy can draw from allowing reusing data at a larger scale. Therefore, all data produced by the project may be published with open access – though this objective will obviously need to be balanced with the other principles described below.

3.2.IPR management and security

As a research and innovation action, VICINITY aims at developing an open framework and gateway – but with support for value added services and business models. The project consortium includes partners from private sector, public sector and end-users. Some partners may have Intellectual Property Rights on their technologies and data. Consequently, the VICINITY consortium will protect that data and crosscheck with the concerned partners before data publication.

A holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity, availability). The security approach will consist of a methodical assessment of security risks followed by their impact analysis. This analysis will be performed on the personal information and data processed by the proposed system, their flows and any risk associated to their processing.

Security measures will include secure protocols (HTTPS and SSL), login procedures, as well as protection about bots and other malicious attacks such as CAPTCHA technologies. Moreover, the industrial demo sites apply monitored and controlled procedures related to the data collection, their integrity and protection. The data protection and privacy of personal information will include protective measures against infiltration as well as physical protection of core parts of the systems and access control measures.

3.3.Personal Data Protection

VICINITY activities will involve human participants as the pilots will be conducted in real apartments and cover real use scenarios related to health monitoring, booking, home management, governance, energy consumption and other various human activity and behaviour analysis –related data gathering purposes. For some of the activities to be carried out by the project, it may be necessary to gather basic personal data (e.g. name, background, contact details, interest, IoT units and assigned actions), even though the project will avoid collecting such data unless data is really necessary for the application.

Such data will be protected in accordance with the EU's Data Protection Directive 95/46/EC[2] "on the protection of individuals with regard to the processing of personal data and on the free movement of such data" and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[2] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en

WP3 and WP4 activities dealing with the implementation and deployment of core components will be performed in Slovakia under leadership of local partners (BVR and IS). For this reason the solution will be reviewed for compliance with Data Protection Act No. 122/2013 approved by National Council of the Slovak Republic together with its amendment No. 84/2014 which already reflects the EC directive proposal 2012/0011/COD.

WP7 and WP8 activities will be performed in Greece, Portugal and Norway under the leadership of local partners. In the following the consortium outlines the legislation for the countries involved in the Trial:

I. Greek Trial in Municipality of Pilea-Hortiatis, Thessaloniki, for Greece, legislation includes "Law 2472/1997 (and its amendment by Law 3471/2006) of the Hellenic Parliament".
   - o ! Regulatory authorities and ethical committees: Hellenic Data Protection Authority http://www.dpa.gr/

II. Norway Trials in Norwegian Helsehus Intelligent Building sites in Halden, Fredrikstad and Lillesand, have to comply with national legislation "Personal Data Act of 14 April No.31" 5relating to the processing of personal data.
   - o ! Each pilot demonstration has to notify regulatory body Datatilsynet pursuant to section 31 of the Personal Data Act and section 29 of the Personal Health Data Filing System Act.

III. Portugal Trial in Martim Longo microgrid pilot site in the Algarve region, Portugal.! The Portuguese renewable energy legislative base dates back to 1988, and was upgraded and reviewed multiple times since then. The most important legislative diplomas are listed; DL 189/88, DL 168/99, DL 312/2001, DL 68/2002, DL 29/2006 and DL 153/2014. The last on the list refers to also one of the most important legislative changes, being the legislative base for broad based auto-consumption, with possibility to inject excess energy in to the grid.
   - o ! The collection and use of personal data in Portugal are regulated by the following two laws: "Law 41/2004" (and its amendment "Law 46/2012"), and "Law 32/2008".

Further information on how personal data collection and handling should be approached in the VICINITY project will be provided in other deliverables.

All personal data collection efforts of the project partners will be established after giving subjects full details on the experiments to be conducted, and obtaining from them a signed informed consent form (see Annex 2: VICINITY consent form template), following the respective guidelines set in VICINITY and as described in section 3.4: Ethics and Security..

Beside this, certain guidelines will be implemented in order to limit the risk of data leaks;

- Keep anonymised data and personal data of respondents separate;
- Encrypt data if it is deemed necessary by the local researchers;
- Store data in at least two separate locations to avoid loss of data;
- Limit the use of USB flash drives;
- Save digital files in one the preferred formats (see Annex 1), and
- Label files in a systematically structured way in order to ensure the coherence of the final dataset

3.4.Ethics and security

The consortium is aware that a number of privacy and data protection issues could be raised by the activities (use case demonstration and evaluation in WP7 and WP8) to be performed in the scope of the project. The project involves the carrying out of data collection in all pilot applications on the virtual neighbourhood. For this reason, human participants will be involved in certain aspects of the project and data will be collected. This will be done in full compliance with any European and national legislation and directives relevant to the country where the data collections are taking place (INTERNATIONAL/EUROPEAN):

- The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data and
- Directive 95/46/EC & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data.

In addition to this, to further ensure that the fundamental human rights and privacy needs of participants are met whilst they take part in the project, in the Evaluation Plans a dedicated section will be delivered for providing ethical and privacy guidelines for the execution of the Industrial Trials. In order to protect the privacy rights of participants, a number of best practice principles will be followed. These include:

- no data will be collected without the explicit informed consent of the individuals under observation. This involves being open with participants about what they are involving themselves in and ensuring that they have agreed fully to the procedures/research being undertaken by giving their explicit consent.
- no data collected will be sold or used for any purposes other than the current project;
- a data minimisation policy will be adopted at all levels of the project and will be supervised by each Industrial Pilot Demonstration responsible. This will ensure that no data which is not strictly necessary to the completion of the current study will be collected;
- any shadow (ancillary) personal data obtained during the course of the research will be immediately cancelled. However, the plan is to minimize this kind of ancillary data as much as possible. Special attention will also be paid to complying with the Council of Europe's Recommendation R(87)15 on the processing of personal data for police purposes, Art.2 :

  > *"The collection of data on individuals solely on the basis that they have a particular racial origin, particular religious convictions, sexual behaviour or political opinions or belong to particular movements or organisations which are not proscribed by law should be prohibited. The collection of data concerning these factors may only be carried out if absolutely necessary for the purposes of a particular inquiry."*

- compensation – if and when provided – will correspond to a simple reimbursement for working hours lost as a result of participating in the study; special attention will be paid to avoid any form of unfair inducement;

- if employees of partner organizations, are to be recruited, specific measures will be in place in order to protect them from a breach of privacy/confidentiality and any potential discrimination; In particular their names will not be made public and their participation will not be communicated to their managers.

3.5.The VICINITY Data Management Portal

VICINITY will develop a data management portal as part of the project. This portal will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been collected and is ready of public distribution. The portal will however not contain any datasets that should not become publicly available.

The initial version of the portal will become available during the 2nd year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available. The VICINITY data management portal will enable project partners to manage and distribute their public datasets through a common infrastructure.

Datasets		Administrative tools
One dataset for each IoT unit	Datasets from pilots (see section 3.5 for examples)	List of sensor / grouping
One dataset for personal information	One dataset for groups of devices	List of actions / sequences
One dataset for energy related domains	One dataset for sequences / actions (combination tokes / nodes)	List of users
• One for each interface	One dataset for each node/object	List of contacts
• One for each measuring device	One dataset for messaging	Balancing loads
• One for each routing device	One dataset for each health device	Booking
One dataset for mobility related domains	One dataset for each smart home device (temperature, smoke, motion, sound)	Messaging
• One for parking data	One dataset for camera	Criteria
• One for booking	One dataset for biometric (fingerprint, retina)	Priorities
• One for areas	One dataset for access	Evaluation / feedback

3.6.Format of datasets

For each data set the following will be specified:

DS. PARTICiPANTName.##.Logical_sensorname	
Data Identification	
Data set description	*Where are the sensor(s) installed? What are they monitoring/registering? What is the dataset comprised of? Will it contain future sub-datasets?*
Source (e.g. which device?)	*How will the dataset be collected? What kind of sensor is being used?*
Partners services and responsibilities	
Partner owner of the device	*What is the name of the owner of the device?*
Partner in charge of the data collection (if different)	*What is the name of the partner in charge of the device? Are there several partners that are cooperating? What are their names?*
Partner in charge of the data analysis (if different)	*The name of the partner.*
Partner in charge of the data storage (if different)	*The name of the partner.*
WPs and tasks	*The data are going to be collected within activities of WPxx and WPxx.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*What is the status with the metadata so far? Has it been defined? What is the content of the metadata (e.g. datatypes like images portraying an action, textual messages, sequences, timestamps etc.)*
Standards, Format, Estimated volume of data	*Has the dataformat been decided on yet? What will it look like?*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Example text:* *Production process recognition and help during the different production phases, avoiding mistakes*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*Example text:* *The full dataset will be confidential and only the members of the consortium will have access on it. Furthermore, if the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination*
Data sharing, re-use and distribution (How?)	*Has the data sharing policies been decided yet? What requirements exists for sharing data? How will the data be shared? Who will decide what to be shared?*
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Who will own the information that has been collected? How will it adhere to partner policies? What kind of limitation are put on the archive?*

3.7.Description of methods for dataset description

Example test dataset will be generated by research teams from the participants in the project. These test datasets will be prepared in XML-files. They will also be made available in XML and JSON format. The datasets will be based on semantic analysis of data from test sensors and applied to an ontology.

The collected dataset will encompass different methodological approaches and IoT standards defined by the global standard initiative oneM2M. The data will run through different test environments like TDD (Test Driven Development), ATDD (Acceptance Test Driven Development), PBT (Property Based Testing), BDD (Behavior Driven Development). The project will focus on using model-based test automation in processes with short release cycles.

Apart from the research teams, these dataset will be useful for other research groups, standardisation organisations and technical integrators with within the area of Internet of Things (IoT).

No comparable data is available as of yet, but there are several descriptions that will be used as basis for the test data.

All datasets are to be shared between the participants during the lifecycle of the project. Feedback from other participants and test implementations will decide when the the dataset should be made publicly available. When the datasets support the framework defined by the VICINITY ontology, they will be made public and presented in open access publications.

The VICINITY partners can use a variety of methods for exploitation and dissemination of the data including:

- Using them in further research activities (outside the action)
- Developing, creating or marketing a product or process
- Creating and providing a service, or
- Using them in standardisation activities

Restrictions: 1) All national reports (which include data and information on the relevant topic) will be available to the public through the HERON web-site or a repository or any other option that the consortium decides and after verification by the partners so as to ensure their quality and credibility.; 2) after month 18 so that partners have the time to produce papers; 3) Open access to the research data itself is not applicable.

3.8.Standards and metadata

The data will be generated and tested through different test automation technologies, e.g. TDL (Test description language), TTCN-3 (Test and Test Control Notation), UTP (UML Testing Profile). The profile should mimic the data communicated from IoT units following the oneM2M specifications.

The modelling language Unified Modelling Language (UML) will be used for the collection, analysis and processing of requirements as well as for the specification message exchanges and overviews of architecture and behaviour specifications.

The project intend to share the datasets in an internally accessible disciplinary repository using descriptive metadata as required/provided by that repository. Additional metadata to example test datasets will be offered within separate XML-files.

They will also be made available in XML and JSON format. Keywords will be added as notations in UML and modelled on the specifications defined by oneM2M. The content will be similar to relevant data from compatible IoT devices and network protocols. No network protocols have been defined yet, but several have been evaluated. Files and folders will be versioned and structured by using a name convention consisting of project name, dataset name, date, version and ID.

3.9.Data sharing

The project aim to prepare the API for internal testing through the VICINITY open gateway.

The VICINITY open gateway is defined as Interoperability as a Service. In other words - it is a cloud based service that assumes the data has already been gathered and transferred to the software running on the service layer. These data will be made available for researchers in a controlled environment, where login credentials is used to get access to the data in XML and JSON-format.

The project focus on developing a framework that that allows for a scalable and futureproof platform upon which it can invest and develop IoT applications, without fear of vendor lock-in or needing to commit to one connectivity technology.

The researchers must therefore be committed to the requirements, architecture, application programming interface (API) specifications, security solutions and mapping to common industry protocols such as CoAP, MQTT and HTTP. Further analysis will be performed using freely available open source software tools. The data will also be made available as separate files.

The goal is to ultimately support the Europe 2020 strategy[3] by offered the open data portal. The Digital Agenda proposes to better exploit the potential of Information and Communication Technologies (ICTs) in order to foster innovation, economic growth and progress. Thus VICINITY will support EUs efforts in exploiting the potential offered by using ICT in areas like climate change, managing ageing population, intelligent transport system with more.

3.10. Archiving and preservation (including storage and backup)

As specified by the "rules of good scientific practice" we aim to preserve data for at least ten years. Approximated end volume of example test dataset is currently 10 GB, but this may be subject to change as the scope of the project may change.

Associated costs for dataset preparation for archiving will be covered by the project itself, while long term preservation will be provided and associated costs covered by a selected disciplinary repository.

During the project data will stored on the VICINITY webcloud as well as being replicated to a separate external server.

[3] https://ec.europa.eu/digital-single-market/en/europe-2020-strategy

4. Datasets for smart grid from Aalborg University (AAU)

AAU will mainly deal with control design, energy management systems implementation and ICT integration in small scale energy systems. AAU will scale-up by using hardware in the loop solution and will participate actively in the implementation at the Energy sites proposed in VICINITY. AAU will act as interface between ICT experts and Energy sites in the project, as well as test interactions between the developed concepts on the ICT side and the control and management of electric power networks. Implementation and experimental results will be an important outcome for the project.

DS.AAU.01.GRID_Status

Data Identification	
Data set description	*This data set comprised different parameters characterising the electrical grid from the generation to the distribution sections. The cost of the electricity will also be considered in this data set, so as to have full information that enables micro-trading actions.*
Source (e.g. which device?)	*The sensors that feed this data set are; Electrical energy generated on-site from RES ,Thermal energy generated on-site, thermal energy consumed, grid electricity consumed, instant grid cost of energy consumed, value of energy purchased from the grid*
Partners services and responsibilities	
Partner owner of the device	*The devices will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*AAU*
Partner in charge of the data analysis (if different)	*AAU*
Partner in charge of the data storage (if different)	*AAU*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at Aalborg University Facilities and AAU servers, allowing access to registered users. Data exploitation is foreseen to be achieved through testing value-added services, allowing full access for authorized personel to data analytics and statistical analysis.*

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized AAU personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a database at the Aalborg University facilities, allowing authorised access to external end-users. A back up will be stored in an external storage device, kept by AAU in a secured place. Data will be kept indefinitely allowing statistical analysis.*

5. Datasets for smart energy from Enercoutim (ENERC)

ENERC will participate providing the facilities and the experience in implementing solar production integrated into municipality smart city efforts. To this end, ENERC will actively participate in the deployment, management and evaluation of the “Smart Energy Microgrid Neighbourhood” Use Case. Its contribution will be focused on the energy resource potential demand studies and economic sustainability. Its expertise will allow ICT integration with smart city management focused on better serving its citizens.

The main aim of this project is the demonstration of a Solar Platform which provides a set of shared infrastructures and reduces the total cost per MW as well as improves the environmental impact compared to the stand alone implementation of these projects. As main responsibilities, ENERC will be in charge of strategic technology planning and integration coordination, designing potential models for municipal energy management, as well as identifying the optimal ownership structure of the microgrid system with a focus on delivering maximum social and economic benefit to the local community.

DS.ENERC.01.METEO_Station	
Data Identification	
Data set description	*The weather conditions will influence the energy production, so it becomes critical to understand the current and foreseen scenarios. Moreover the control of those parameters*
Source (e.g. which device?)	*The sensors that feed this data set are; temperature, humidity, wind speed and wind direction, barometer, precipitation measurement and sun tracker*
Partners services and responsibilities	
Partner owner of the device	*The devices will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*ENERC*
Partner in charge of the data analysis (if different)	*ENERC*
Partner in charge of the data storage (if different)	*ENERC*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.*
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized ENERC personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.*

DS.ENERC.02.BUILDING_Status

Data Identification	
Data set description	*The information associated to the energy consumption in buildings will allow identifying the usage of resources for each measurement point.*
Source (e.g. which device?)	*The sensors that feed this data set are; Cooling energy demand, heating energy demand, hot water demand, building equipment demand*
Partners services and responsibilities	
Partner owner of the device	*The devices will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*ENERC*
Partner in charge of the data analysis (if different)	*ENERC*
Partner in charge of the data storage (if different)	*ENERC*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	

Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized ENERC personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.*

DS.ENERC.03.GRID_Status

Data Identification	
Data set description	*This data set comprises the different parameters that characterise the electrical grid from the generation to the distribution sections. Moreover the cost of the electricity will be consider in this data set so as to have full information that enables micro-trading actions.*
Source (e.g. which device?)	*The sensors that feed this data set are; Electrical energy generated on-site from RES ,Thermal energy generated on-site, thermal energy consumed, grid electricity consumed, instant grid cost of energy consumed, value of energy purchased from the grid*
Partners services and responsibilities	
Partner owner of the device	*The devices will be the property of the test site owners, where the data collection is going to be performed.*

Partner in charge of the data collection (if different)	*ENERC, AAL*
Partner in charge of the data analysis (if different)	*ENERC, AAL*
Partner in charge of the data storage (if different)	*ENERC, AAL*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged devices. However, it is estimated to be 4 KB/transmission.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at SOLAR LAB Facilities and AAL servers, allowing access to registered users. Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. facility managers), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized ENERC/AAL personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the SOLAR LAB facilities and AAL servers, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by ENERC in a secured place. Data will be kept indefinitely allowing statistical analysis.*

6. Datasets for eHealth from GNOMON Informatics SA (GNOMON)

GNOMON will provide its background knowledge in the specific field of assisted living and tele care in the context of social workers. In addition, GNOMON will actively contribute in the use case pilot setup, assessment and benchmarking.

The company has developed and provided the remote care and monitoring integrated system for people with heath problems as well as of the software applications for support and organization using information and communication technologies of the business operation of HELP AT HOME program in the Municipality of Pilea-Hortiatis . This infrastructure could be further exploited and extended for the scope of VICINITY project and specifically for the realisation of the eHealth Use Case.

DS.GNOMON.01.Pressure_sensor

Data Identification	
Data set description	*The sensors will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The measurements are scheduled to be taken once a day, requiring the patient to make use of the device placed within their apartment. The main task of the sensor is to monitor pressure (systolic/diastolic) and heart rate levels.*
Source (e.g. which device?)	*The dataset will be collected via a combination of connected devices consisting of a Bluetooth Blood Pressure monitor and a Connectivity Gateway based on Raspberry pi.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*GNOMON, MPH*
Partner in charge of the data analysis (if different)	*GNOMON, MPH*
Partner in charge of the data storage (if different)	*GNOMON, MPH*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in XML format. In a later stage, they are converted to JSON format and stored in a database. Regarding the volume of data, it depends on the participation levels of the engaged patients. However, it is estimated to be 16 KB/measurement.*
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. patient's doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.*

DS.GNOMON.02.Weight_sensor

Data Identification	
Data set description	*The sensors will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The measurements are scheduled to be taken once a day, requiring the patient to make use of the device placed within their apartment. The main task of the sensor is to keep track of weight measurements and mass index (given the fact that the patient provides an accurate value of his/her height). Future subset may contain information about resting metabolism, visceral fat level, skeletal muscle and body age.*
Source (e.g. which device?)	*The dataset will be collected via a combination of connected devices consisting of a Bluetooth Body Composition monitor and a Connectivity Gateway based on Raspberry pi.*
Partners services and responsibilities	

Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*GNOMON, MPH*
Partner in charge of the data analysis (if different)	*GNOMON, MPH*
Partner in charge of the data storage (if different)	*GNOMON, MPH*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in XML format. In a later stage, they are converted to JSON format and stored in a database. Regarding the volume of data, it depends on the participation levels of the engaged patients. However, it is estimated to be 48 KB/measurement.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. patient's doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.*

DS.GNOMON.03.Fall_sensor

Data Identification	
Data set description	*The fall sensors is a wearable sensor that will be in possession of patients in need of assisted living and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The main goal of the sensor is to automatically detect when a patient falls either due to an accident or in the case of a medical incident. The event is triggered automatically after a fall, but a similar event is also triggered by pressing the equivalent panic button (wearable actuator). In both cases, an automated emergency phone call is placed to the eHealth Call Center.*
Source (e.g. which device?)	*The dataset will be collected via a combination of devices consisting of a hub (Lifeline Vi*[4]*) and a fall detector that are wirelessly connected.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*GNOMON, MPH*
Partner in charge of the data analysis (if different)	*GNOMON, MPH*
Partner in charge of the data storage (if different)	*GNOMON, MPH*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*An audit log containing alerts (incl. false alarms) is stored. The amount of alerts is estimated to be 50 alerts (incl. false alarms) per month.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. patient's doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed*

[4] http://www.tunstall.co.uk/solutions/lifeline-vi

	confidentiality framework. *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.*

DS.GNOMON.04.GPS_Sensor

Data Identification	
Data set description	*The GPS sensor is a wearable device that will be in possession of patients in need of assisted living (mainly suffering from dementia) and identified by the equivalent municipality (MPH) health care services to ensure the validity of each case. The main task of the sensor is to provide the subject's coordinates, in order to prevent his/her disappearance.*
Source (e.g. which device?)	*The dataset collected by the sensor will be transmitted through GSM.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed.*
Partner in charge of the data collection (if different)	*GNOMON, MPH*
Partner in charge of the data analysis (if different)	*GNOMON, MPH*
Partner in charge of the data storage (if different)	*GNOMON, MPH*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata may include device id, measurement date, device owner, state of the monitored activity, etc.*
Standards, Format, Estimated volume of data	*The data are received in JSON format. Regarding the volume of data, it depends on the motion/activity levels of the engaged patients. However, it is estimated to be 4 KB/transmission.*
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	*Due to privacy issues, the collected data are stored at a secured database scheme at MPH headquarters, allowing access to registered users (i.e. MPH health care services personnel and eHealth call center). Data exploitation is foreseen to be extended through envisioned value-added services, allowing full access to specific authorised users (e.g. patient's doctors), and for a broader use in an anonymised/aggregated manner for data analytics and statistical analysis.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the authorized MPH personnel and related end-users will have access as defined. Specific consortium members involved in technical development and pilot deployment will further have access under a detailed confidentiality framework.* *Furthermore, if the dataset in an anonymised/aggregated manner is decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Due to ethical and privacy issues, data will be stored in a database scheme at the headquarters of MPH, allowing only authorised access to external end-users. A back up will be stored in an external storage device, kept by MPH in a secured place. Data will be kept indefinitely allowing statistical analysis.*

7. Datasets for intelligent mobility from Hafenstrom AS (HITS)

HITS will be responsible for the user requirements specifications and demonstration of transport domain use case, while it will actively participate in the dissemination and exploitation activities of the project.

HITS will be responsible for the Use cases “Virtual Neighbourhood of Buildings for Assisted Living integrated in a Smart Grid Energy Ecosystem” and “Virtual Neighbourhood of Intelligent (Transport) Parking Space”. Towards this direction, it will be the main partner to bring/arrange the required infrastructure, in collaboration with other Consortium partners (i.e., TINYM partner), for the use case demonstration.

DS.HITS.01.Parkingsensor

Data Identification	
Data set description	*The sensors will be installed at a testsite, and will register proximity of objects of a certain size. Future subset may contain information about temperature, humidity, noise, light and other temperature, visual and touch related data.The sensors main task is to detect if the space is occupied. This information will later on be integrated with identifaction in order to verify that the vehicle/unit that occupies the space is licenced through either booking or ticketing action being taken.*
Source (e.g. which device?)	*The dataset will be collected through a sensor that is mounted at the parking site.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed*
Partner in charge of the data collection (if different)	*HITS*
Partner in charge of the data analysis (if different)	*HITS*
Partner in charge of the data storage (if different)	*HITS*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.*
Standards, Format, Estimated volume of data	*The data will be stored at XML format and are estimated to be 50-300 MB per month.*
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	*Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be available to participants in the project. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device.*

DS.HITS.02.SmartLighting

Data Identification	
Data set description	*Smartlighting will be installed at the lab, and will demonstrate how light and colours can indicate the state of access and availability. Future subset may contain information about proximity, movement, heat sensing (infrared), sound sensing and door contact sensors. The smart lights main task is to visually inform about the state of the parking space. This information may later on be integrated with indicators for occupancy, time to availability and validity.*
Source (e.g. which device?)	*The dataset will be received from a laptop in the lab.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed*
Partner in charge of the data collection (if different)	*HITS*
Partner in charge of the data analysis (if different)	*HITS*
Partner in charge of the data storage (if different)	*HITS*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.*

Standards, Format, Estimated volume of data	*The data will be stored at XML format and are estimated to be 50-300 MB per month.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be available to the members of the consortium. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device.*

DS.HITS.03.LaptopTeststation

Data Identification	
Data set description	*The laptop test station will be installed at the workbench where the operator normally works, and will aggregate data and process information received wirelessly from other devices delivering data of relevance to the mobility domain and parking in particular. Future subset may contain information about other domains – energy, and datapackages from smart home and health-devices. The test stations main task is to process data and trigger activate and log actions accordingly.*
Source (e.g. which device?)	*The dataset will be collected wirelessly and via USB ports.*
Partners services and responsibilities	
Partner owner of the device	*The device will be the property of the test site owners, where the data collection is going to be performed*
Partner in charge of the data collection (if different)	*HITS*
Partner in charge of the data analysis (if different)	*HITS*
Partner in charge of the data storage (if different)	*HITS*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	

Info about metadata (Production and storage dates, places) and documentation?	*The dataset will be accompanied with the respective documentation of its contents. Indicative metadata include: (a) description of the experimental setup (e.g. process system, date, etc.) and procedure which is related to the dataset (e.g. proactive maintenance action, unplanned event, nominal operation. etc.), (b) scenario related procedures, state of the monitored activity and involved workers, involved system etc.*
Standards, Format, Estimated volume of data	*The data will be stored at XML format and are estimated to be 50-300 MB per month.*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Registering parking activity based upon availability, vehicle, ownership/licence, comparing with nearby infrastructure and surrounding ITS technology.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be available to the members of the consortium. If the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, a data management portal will be created that should provide a description of the dataset and link to a download section. Of course these data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination.*
Data sharing, re-use and distribution (How?)	*The created dataset could be shared by using open APIs through the middleware as well as a data management portal.*
Embargo periods (if any)	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in the storage device of the developed system (computer). A back up will be stored in an external storage device.*

8. Datasets for smart buildings from Tiny Mesh AS (TINYM)

The primary role of Tiny Mesh Company is as a developer and technology provider, with the company´s IoT solution as the main enabling technology. The focus is on participate in use case with technology that solutions that offer the promise to create new products, services and business model by connecting, integrating and controlling all kinds of meters, street lights, sensors, actuators, assets, devices, tags and other things as part of the Internet-of-Everything.

TINYM will contribute in the practical implementation through their work with definitions of use case Assisted Living. TINYM will take practical ownership of the various demo sites through the role as of leader of WP7.

DS. TinyMesh.01.IEQ_Sensor

Data Identification	
Data set description	*The sensors will be installed at the in room in buildings where there is a need for monitoring of indoor climate. Data packet contains sensor data such as temperature, humidity, light (lux), sound pressure (db), CO2 and movement.* *This information is input data for analysis of building performance, use load and energy efficiency, which in turn forms the basis for estimates of the building's energy flexibility*
Source (e.g. which device?)	*The IEQ Sensor has a mesh based communication module and put the sensor into a member of a Tinymesh Network. The Tinymesh network deliver all the data packed from sensors in the net to one or more Tinymesh Gateways.*
Partners services and responsibilities	
Partner owner of the device	*The property owner ore the producer of the IEQ Sensor (Serinus Technology)*
Partner in charge of the data collection (if different)	*TinyMesh*
Partner in charge of the data analysis (if different)	*Tiny Mesh*
Partner in charge of the data storage (if different)	*Tiny Mesh*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata about location of the sensor, network topology and network status will be available in Tinymesh Workbench.*
Standards, Format, Estimated volume of data	*The information element will be defined by the Building Smart Data Dictionary. (BSDD) The payload from one device is 120 kB. Datavolum deepens of the sampling frequency in the sensor.*
Data exploitation and sharing	

Data exploitation (purpose/use of the data analysis)	*The purpose of this collection is to give input data for analysis of building performance, use load and energy efficiency, which in turn forms the basis for estimates of the building's energy flexibility*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the members of the consortium will have access on it. Furthermore, if the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, an API will be an interface to the dataset. Data will be anonymized, so as not to have any potential ethical issues with their publication and dissemination*
Data sharing, re-use and distribution (How?)	*The full dataset will be confidential and only the members of the consortium will have access on it for privacy reasons.*
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored applications to Tiny Mesh and Serinus technology.*

DS. TinyMesh.02 Tinymesh_Gateway

Data Identification	
Data set description	*Datapacked form any Tinymesh network*
Source (e.g. which device?)	*Tinymesh Gateway in a transparent communicationgateway to connect any Tinymesh network to the Tinymesh Cloud service.*
Partners services and responsibilities	
Partner owner of the device	*Tiny Mesh AS*
Partner in charge of the data collection (if different)	*Tiny Mesh AS*
Partner in charge of the data analysis (if different)	*Tiny Mesh AS*
Partner in charge of the data storage (if different)	*Tiny Mesh AS*
WPs and tasks	*The data are going to be collected within activities of WP7 and WP8.*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Tinymesh Gateway is a wireless transparent serial communication device that can transfer data in two modus; transparent and packed.*
Standards, Format, Estimated volume of data	-
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*Tinymesh Gateway is a wireless transparent serial communication device*

Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the members of the consortium will have access on it.*
Data sharing, re-use and distribution (How?)	*Data and metadata will be accessible by an API in Tinymesh Cloud ore in IEQ Audit form Serinus Technology.*
Embargo periods (if any)	-
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data and metadata will be accessible by an API in Tinymesh Cloud ore in IEQ Audit form Serinus Technology.*

9. Conclusions

The VICINITY DMP is based upon the datasets for procedures and infrastructure that are anticipated at this point in the project. The next deliverable is set to M24 (December 2017), but in order to assist the official project review process by the commission for the first project period (M1-M24), a preliminary version of the updated DMP of D9.3 can be further delivered prior to M24 (December 2017), in order to be enable a better assessment of the progress of the Data Management in the project by the reviewers.

The next actions will be to focus on semantics and further clarification of procedures, participant and stakeholder engagement and identifying areas that need special attention. Activities for a Data Management Portal will have been initiated, and changes to the datasets may have been made after on more comprehensive studies of the pilot sites have been conducted. Consent forms will have been adapted and signed by involved stakeholders, and the necessary infrastructure will be set up based on the feedback from involved parties.

Lessons learned from this report is that nearly all project partners will be owners or/and producers of data. This means that a Data Management Portal will need to allow for specific access to each project partner, and that editing / access rights will need to be managed accordingly. It must also be noted that the partners are unable to exactly specify what kind of datasets that will be relevant as the project proceeds. This is what they expect to learn from the pilot sites and other tests conducted at the workbench. It is therefore expected that the datasets may change accordingly.

The VICINITY Data Management Plan will put a strong emphasis of the appropriate collection – and publication should the data be published – of metadata, storing all the information necessary for the optimal use and reuse of those datasets. This metadata will be managed by each data producer, and will be integrated in the Data Management Portal.

Once the solutions to be tested by the project and the content of this testing are specified, the next step – which will be described in the updated version of this report due in December 2017 – will be to finalize the specifications of the Data Management Portal of the project and provide information on the existence (or not) of similar data and the possibilities for integration and reuse. In addition, issues like the period of data preservation, the approximated end volume, the associated costs and how these are planned to be covered will be tackled in order to make the Portal and other necessary management tools operational and to provide a detailed Management Plan for each data set.

References

European Commission. (2013). *Guidelines on Data Management in Horizon 2020*. Retrieved 2 June, 2015, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European IPR Helpdesk. (2014). *Fact Sheet Open Access to publications and data in Horizon 2020: Frequently Asked Questions (FAQ)*. Retrieved 3 July, 2015, from https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf

Annex 1 – Preferred Formats

1.1.Selection of File Formats

All formats of digital files stand the risk of becoming obsolete in the future. If a file format becomes obsolete, it means that the current software will not be able to represent and use the content of the file in the way it was meant to at the time of creation.

To prevent file format obsolescence, some precautions can be taken. One such measure is to select file formats which have a high chance of remaining usable in the far future. As a general guideline, VICINITY considers that the file formats best suited for long-time preservation and accessibility:

1. are commonly used;
2. have open specifications;
3. are independent of specific software, developers or suppliers.

However, it is not always possible to select formats that meet with all of these ideal attributes.

1.2.Preferred and Acceptable Formats

VICINIY has assessed a number of file formats resulting in a list of preferred formats and acceptable formats. This list will change over time as new formats will be developed and others will fall into disuse.

Preferred formats are the file formats which can be trusted to offer the best long-term guarantees for usability, accessibility and robustness. In principle, VICINITY expects these formats to be durable for the long term. VICINITY will accept research data deposited in preferred formats in VICINITYs repository without question.

Acceptable formats are file formats which are commonly used besides the preferred formats; have average to reasonable scores regarding their usability, accessibility and robustness in the long term. VICINITY strongly prefers the use of preferred formats but in most cases, the use of acceptable formats will be allowed in to the archive as well.

Table 1 presents a summarised overview of VICINITYs preferred and acceptable Formats.

Table 1 – Preferred and acceptable formats for data storage

	Preferred format(s)	Acceptable format(s)
Text documents	PDF/A (.pdf)	OpenDocument Text (.odt) MS Word (.doc, .docx) Rich Text File (.rtf) PDF (.pdf)
Text file	Unicode TXT (.txt, ...)	Non-Unicode TXT (.txt, ...)
Marked-up language		XML (.xml) HTML (.html)

	Preferred format(s)	**Acceptable format(s)**
Spreadsheets	OpenDocument Spreadsheet (.ods) Comma Separated Values (.csv)	MS Excel (.xls, .xlsx) PDF/A (.pdf) OOXML (.docx, .docm)
Databases	ANSI SQL (.sql, …) Comma Separated Values (.csv)	MS Access (.mdb, .accdb) dBase III or IV (.dbf)
Statistical data	R SPSS Portable (.por) SAS transport (.sas) STATA (.dta)	
Images (raster)	JPEG (.jpg, .jpeg) TIFF (.tif, .tiff) PNG (.png)	JPEG 2000 (.jp2)
Images (vector)	Scalable Vector Graphics (.svg)	Adobe Illustrator (.ai) PostScript (.eps)
Video	MPEG-2 (.mpg, .mpeg, …) MPEG-4 H264 (.mp4) Lossless AVI (.avi) QuickTime (.mov)	
Audio	WAVE (.wav)	MP3 AAC (.mp3)
Computer Aided Design (CAD)	AutoCAD DXF versie R12 (.dxf)	AutoCAD other versions (.dwg, .dxf)
Geographic Information (GIS)	Geographic Markup Language (.gml) MapInfo Interchange Fomat (.mif/.mid)	ESRI Shapefiles (.shp and associated files) MapInfo (.tab and associated files) Keyhole Markup Language (.kml)
Images (georeferenced)	GeoTIFF (.tif, .tiff)	TIFF World File (.tfw en .tif)
Raster GIS	ASCII GRID (.asc, .txt)	ESRI GRID (.grd and associated files)
3D	WaveFront Object (.obj) X3D (.x3d)	COLLADA (.dae) Autodesk FBX (.fbx)
RDF	W3C standards	

Annex 2: VICINITY Consent form Template

A template of the consent form to be used is presented below, to be adopted as required per pilot use case.

VICINITY
2020

Consent Form

Purpose of the study

A commonly understandable written description of the project and its goals (2-3 paragraphs)

Planned Project Progress

The planned project progress and the related testing and evaluation procedures (1-2 paragraphs)

Disclaimer Rights

Advice on unrestricted disclaimer rights on their agreement.

Voluntary Participation Form for the needs of the VICINITY project

1. Participant Information

Basic information and participant's reference code ID (the reference code ID will be used throughout the pilot trial execution)

2. Study Information

Details about the pilot Use Case

3. Participant's Questionnaire

Questions verifying that the participant:

- has been fully informed on the purpose, duration, procedures of the study;

- has been informed on the rights to deny participating or to quit from the study and about the corresponding consequences.

- has been informed on the contact person in case that I have questions and queries about the study.

- had adequate time to make my decision concerning my participation in the study.

- comprehend that he/she can quit from the study at any time without having to justify his/her decision.

- has been informed about potential effects, difficulties and dangers.

- has been informed about the sensors equipment that will be used to collect data.

- has been informed about the security of the study data and results.

- has been ensured about the confidentiality of his/her personal information. Publications of the study results do not allow the personal data recognition, due to the principle of anonymity. Always under the confidentiality principles.

4. Signed Consent to Participate

A signed consent of the participant allowing the study responsible to examine and inspect the data collected during the study.